Computers and Electrical Engineering 000 (2017) 1–13

Contents lists available at ScienceDirect

Computers and Electrical Engineering

journal homepage: www.elsevier.com/locate/compeleceng

Pattern mining-based video saliency detection: Application to moving object segmentation☆

Hiba Ramadan*, Hamid Tairi

LIIAN Laboratory, Department of Informatics Faculty of Sciences Dhar-Mahraz, University of Sidi Mohamed Ben Abdellah, Fez, Morocco

ARTICLE INFO

Article history:
Received 28 February 2017
Revised 27 August 2017
Accepted 28 August 2017
Available online xxx

Keywords:
Spatiotemporal saliency
Pattern mining algorithm
Spatiotemporal saliency patterns
Image saliency
Motion saliency
Moving object segmentation
Energy minimization

ABSTRACT

In this paper, we present a new model for spatiotemporal saliency detection. Instead of previous works which combine the image saliency in the spatial domain with motion cues to build their video saliency model, we propose to apply the pattern mining (PM) algorithm. From initial saliency maps computed in spatial and temporal domains, discriminative spatiotemporal saliency patterns can be recognized and their label information is propagated to obtain the final saliency map. Our model ensures a good compromise between image saliency and motion saliency and presents an accurate prediction to estimate salient regions in comparison with other methods for video saliency detection. Finally, as an application of our algorithm, our spatiotemporal saliency map is combined with appearance models and dynamic location models into an energy minimization framework to segment salient moving object. Experiments show a good performance of our algorithm for moving object segmentation (MOS) on benchmark datasets.

1. Introduction

Recently, the emulation of human cognitive ability to distinct the most attractive regions in image and video is one of the most challenging and progressing trends in computer vision, due to its wide applications such as object detection/segmentation/recognition, content-based image/video compression, visual tracking and many more [1]. The methods of saliency detection can be divided into two categories: bottom-up or top-down approaches. The bottom-up models exploit the features extracted from the image for saliency computation such as intensity, color and orientation. The top-down approaches require an explicit understanding of the context of the image: supervised learning with a specific class is the most adopted principle in this category. A good review of saliency detection models, its applications and classification of them can be found in [1].

In this work, we focus on video saliency detection, where the goal is to identify the most salient regions in space and time by integrating motion cues in addition to the spatial information. Differently to the majority of existing methods which combine the image saliency with motion features [2–6], we propose in this paper a new model for video saliency detection using the pattern (PM) algorithm. Inspired by the work of [7] proposed to promote existing image saliency methods, we extend its principle into the spatiotemporal domain to estimate salient regions in video. First, for each frame from the input video, we compute two saliency maps: spatial saliency and temporal saliency. Then, from the generated maps, initial background and foreground regions can be detected and used as input of the PM algorithm to identify spatiotemporal saliency

☆ Reviews processed and recommended for publication to the Editor-in-Chief by Area Editor Dr. E. Cabal-Yepez.

* Corresponding author.
E-mail addresses: hiba.ramadan@usmba.ac.ma (H. Ramadan), hamid.tairi@usmba.ac.ma (H. Tairi).

http://dx.doi.org/10.1016/j.compeleceng.2017.08.029
0045-7906/© 2017 Elsevier Ltd. All rights reserved.

patterns. Finally, a label propagation process is applied to obtain the spatiotemporal saliency map. To illustrate the efficiency of our algorithm, we integrate the resultant saliency map into an energy minimization framework to segment salient moving object.

This paper is organized as follows: Section 2 presents the related work including video saliency detection and moving object segmentation. In Section 3 we describe the different steps of our proposed algorithm. Finally, experiments and comparative study are given in Section 4.

2. Related work

2.1. Video saliency detection

The most of video saliency models proposed in the literature combine the image saliency in the spatial domain with motion cues. In [2], Gao et al., added the motion channel for prediction of human eye fixations in dynamic scenes based on the center-surround hypothesis, to the image saliency. Ejaz et al. [3] proposed a feature aggregation based visual attention model combined with motion intensity for video summarization. Recently, a spatiotemporal saliency model based on superpixel-level trajectories (SLT) by integrating a pixel-level temporal saliency map with a pixel-level spatial saliency map is presented in [4]. In [5], the spatial and temporal saliency maps are merged into one, using a spatiotemporal adaptive entropy-based uncertainty weighting (SCW) approach. The work of [6] integrated both of static edges and motion boundaries to build a spatiotemporal edge map, then geodesic distance is used to estimate the final spatiotemporal saliency map, so called geodesic saliency (GS). Singh et al. [8] presented a learning based method (LTSPS) that computes video saliency by integrating color dissimilarity, objectness measure, motion difference, and boundary score.

2.2. Moving object segmentation

The use of saliency models to segment moving object in video is one of the most promising approaches [7,9–12]. Fukuchi et al. [9] exploited visual saliency priors via the maximum a posteriori estimation of the Markov random field to perform MOS. Li et al. [10] proposed a video object extraction framework by applying a conditional random field to combine visual and motion saliency cues. The work of [7] achieved a spatiotemporal object segmentation based on the proposed GS combined with appearance and location models, and then the MOS is solved as a pixel labelling problem using energy minimization. In [11], a spatiotemporal-based approach is proposed to segment both static and dynamic objects by using motion and image saliency. In [12], the authors propose a spatiotemporal visual saliency map for MOS, by combining the gradient magnitude of optical flow and the visual saliency to extract a moving region of interest which serves as seeds for MOS using the convex active contour.

In addition to these saliency based methods, there are also other techniques that have been proposed for MOS based on tracking keypoints over time [13,14], using general background subtraction models for moving camera [15], based on object proposals [16,17] or appearance modeling based techniques [18,19].

3. Proposed algorithm

Our main contribution in this work consists of using the PM algorithm to build our spatiotemporal saliency map instead of simple combination between spatial and motion cues as previous works. First, for each frame from the input video sequence, we apply the de-texturing method of [20] to preserve only the cartoon component of the image without unwanted edges and textures. Then the saliency map of the de-textured image in the spatial domain is obtained using [21]. In the other hand, we compute the optical flow [22] of the current frame to extract its saliency flow. In addition to the saliency maps computed in the spatial and the temporal domains, we use temporal superpixels to generate input regions of the PM algorithm. According to the mined saliency patterns, pertinent background and foreground seeds can be detected and their label information is propagated to generate the spatiotemporal saliency map. Finally, the MOS is performed by integrating the obtained saliency map with appearance and location models into an energy minimization framework to obtain the final labeling for the processed frame (Fig. 1).

3.1. Pre-processing: structure image extraction

Before saliency generation, we apply to each frame from the input video sequence the structure extraction algorithm [20] for pre-processing. In fact, this de-texturing procedure smoothens the local gradients in textures, preserves the global structures of the objects, diminishes insignificant details and gives more accurate boundary information. The structure extraction from texture via relative total variation (STRTV) proposed in [20] is a simple and effective method based on local variation measures to accomplish texture removal; it uses fast and robust numerical solver and can process non-uniform and anisotropic texture. The objective function is given by:

$$argmin_S \sum_p (St_p - I_p) + \lambda \cdot \left(\frac{D_x(p)}{L_x(p) + \varepsilon} + \frac{D_y(p)}{L_y(p) + \varepsilon} \right) \quad (1)$$

Please cite this article as: H. Ramadan, H. Tairi, Pattern mining-based video saliency detection: Application to moving object segmentation, Computers and Electrical Engineering (2017), http://dx.doi.org/10.1016/j.compeleceng.2017.08.029